

Multimedia Accessibility Quick Fix Guide¹

Our guide provides simple, actionable steps to ensure your recorded video and audio content is accessible to all through captions, transcripts, and audio descriptions with minimal time investments.

Vocabulary

- **Caption:** Synchronized text of all speech and relevant sounds, displayed on-screen in real-time with video, include sound descriptions (e.g., music playing, audience laughter)
- **Transcript:** A text-based, word-for-word record of a video's spoken dialogue, audio, and non-speech sounds that accompanies a video and is accessed separately from the video.
- **Descriptive Transcript:** Used when videos do not have sound, a descriptive transcript is a written description of all visual content that includes what is happening on screen, visual changes, and text displayed. This is a good alternative when audio description is not feasible.
- **Audio description:** A narration of key visual elements in video, describing actions, settings, facial expressions, and on-screen text during natural pauses in dialogue.

Path 1: Audio-Only Content (e.g., podcasts, audio recordings, voice memos, audio lectures)

- **You Need:** Transcript
- **Include:** Speaker names/roles, all dialogue and speech content, descriptions of non-speech sounds (music, laughter, applause), and relevant background sounds that provide context

Path 2: Video with Audio (e.g., recorded lectures, interviews, demonstrations, presentations)

- **You Need:** Captions and transcript
- **Pro-Tip for Future Videos:** Describe visual content as you record so that you don't need to later add audio description.

Path 3: Video-Only or Silent Video (e.g., visual demonstrations with no narration)

- **You Need:** Audio description or descriptive transcript
- **Include:** Descriptions of visual actions, on-screen text, scene changes, and demonstrations

¹This document is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International license](https://creativecommons.org/licenses/by-nc-sa/4.0/) (CC BY-NC-SA 4.0 International). Users are welcome to distribute, adapt, remix, and build upon the material in any medium or format if they provide attribution to USM Kirwan Center for Academic Innovation and any other identified original authors, provide a link to the Creative Commons license, and indicate if changes were made. Users may not use the material for [commercial purposes](#), and if users remix, transform, or build upon the material, they must distribute their contributions under the same CC BY-NC-SA license.

Tools to Support Multimedia Remediation

- Zoom, Panopto, and YouTube auto-generated captions (require human review)
- Vimeo auto-generated captions on paid plans (require human review)
- Panopto and Zoom audio description available
- YouDescribe is a free web tool for adding audio descriptions to YouTube videos

Editing Essentials

Technical Language & Discipline-Specific Terms

- **The Problem:** AI auto-captioning struggles with specialized vocabulary, acronyms, formulas, and proper nouns
- **The Solution:** Create a list of key terms used in your video; use “find/replace” in your caption editor to fix recurring errors quickly

Non-Native Speakers & Cognitive Load

Consider “Clean Captions”:

- Remove excessive filler words (um, uh, like, you know)
- Fix false starts and incomplete sentences
- Maintain natural speech rhythm while improving readability
- Don’t over-edit; preserve authentic communication style

Accuracy Standard

- The industry benchmark for caption accuracy is 99%, meaning minimal errors in transcription
- As of Spring 2026, standard AI-generated auto-captioning does not meet this benchmark on its own.
- Critical areas requiring human review include:
 - Proper nouns (people, places, organizations)
 - Technical terminology and jargon
 - Context-specific language (i.e., homophones)
 - Numbers, dates, measurements
 - Speaker identification

Sound Descriptions Matter

- Include meaningful non-speech audio (upbeat music playing; students laughing; door slams)
- Why it matters: Context that hearing users get automatically must be made explicit for Deaf/hard of hearing users